

**METHOD AND APPARATUS FOR THE AUTOMATIC SEPARATING
AND INDEXING OF MULTI-SPEAKER CONVERSATIONS**

Abstract

Disclosed are a method and apparatus for processing a continuous audio stream
5 containing human speech in order to locate a particular speech-based transaction in the
audio stream, applying both known speaker recognition and speech recognition
techniques. Hereby it is enabled that only the utterances of a particular predetermined
speaker are transcribed thus providing an index and a summary of the underlying
dialogue(s).

10 In a first scenario, an incoming audio stream, e.g. a speech call from outside, is
scanned in order to detect audio segments of the predetermined speaker. These audio
segments are then indexed and only the indexed segments are transcribed into spoken or
written language. Thus an already occurred specific transaction can be found on an
endless storage media like a magnetic tape. The proposed mechanism thus makes the
15 task of locating an audio log of a specific transaction a much more less effort.

In a second scenario, two or more speakers located in one room are using a
multi-user speech recognition system (SRS). For each user there exists, a different

speaker model and optionally a different dictionary or vocabulary of words already known or trained by the speech or voice recognition system. In such an environment, the invention allows to switch between different dictionaries when a first user has stopped utterance and a second user is going to start his utterance.

5 (Fig. 1B)